13 On Two Types of GA-Learning

Nicolaas J. Vriend

Queen Mary, University of London, UK
n.vriend@qmw.ac.uk

Abstract. We distinguish two types of learning with a Genetic Algorithm. A population learning Genetic Algorithm (or pure GA), and an individual learning Genetic Algorithm (basically a GA combined with a Classifier System). The difference between these two types of GA is often neglected, but we show that for a broad class of problems this difference is essential as it may lead to widely differing performances. The underlying cause for this is a so called spite effect.

13.1 Introduction

One dimension in which we can distinguish types of Genetic Algorithms is the level at which learning is modeled. A first possibility is as a model of *population learning*, which might be called a pure Genetic Algorithm. There is a population of rules, with each rule specifying some action. The fitness of the rules is determined by all members of the population executing their specified action, and observing the thus generated feedback from the environment. On the basis of these performances, the population of rules is modified by applying some genetic operators . The second way to implement a GA is to use it as a model of *individual learning* , which is basically a Classifier System with on top of it a Genetic Algorithm.

The difference between these two types of GA is often neglected, but we show that for a broad class of problems this difference is essential. [1] This difference is due to the fact that when a GA is learning, there are actually two processes going on. On the one hand, when the rules are executed they interact with each other in the environment, generating the outcomes that determine the fitness of each rule. On the other hand, given the thus generated fitnesses of the rules, there is the learning process as such. As we will make precise below, it is the way in which these processes interact with each other that causes the essential difference between the two versions of a GA.

The phenomenon causing this essential difference between an individual and a GA is called the "spite effect". The spite effect occurs when choosing an action that hurts oneself but others even more. [2] In order to introduce the

[1] A more detailed analysis of the wider issue of the essential difference between individual and social learning and the consequence for computational analyses in a more general sense can be found in [12].

[2] The term "spiteful behavior" goes back, at least, to [3]. It has been examined in the economics (see, e.g., [9]), experimental economics (see, e.g., [2], or [4]), and evolutionary game theory literature (see, e.g., [6], [7], and [11]).

essence of the spite effect, consider the bimatrix game in Fig. 13.1 (see [6]), where T and B are the two possible strategies, and a, b, c, and d are the payoffs to the row and column player, with $a > b > c > d$. Clearly, (T, T) is the only Nash equilibrium since no player can improve by deviating from it, and this is the only combination for which this holds. Now, consider the strategy pair (B, T), leading to the payoffs (b, c). Remember that $a > b > c > d$. Hence, by deviating from the Nash equilibrium, the row player hurts herself, but she hurts the column player even more. We could also look at it from the other side. Suppose both players are currently playing strategy B, when the column player, for one reason or another, deviates to play strategy T instead, thereby improving his payoff from d to c. But the row player, simply sticking to her strategy B, would be "free riding" from the same payoff of d to a payoff c that is even higher than b. The question, then, addressed in this paper is how this spite effect influences the outcomes of an individual learning GA and a population learning GA.

	T	B
T	a, a	c, b
B	b, c	d, d

Fig. 13.1. Bimatrix game, with payoffs $a > b > c > d$

The remainder of this paper is organized as follows. In Section 13.2 we present an example illustrating this difference, which we will analyze in Section 13.3 in relation to the spite effect. Section 13.4 draws our example into a broader perspective by discussing some specific features of the example, and presents some conclusions.

13.2 An Example

Consider a standard Cournot oligopoly game. There is a number n of firms producing the same homogeneous commodity, who compete all in the same market. The only decision variable for firm i is the quantity q_i to be produced. Once production has taken place, for all firms simultaneously, the firms bring their output to the market, where the market price P is determined through the confrontation of market demand and supply. Let us assume that the inverse demand function is $P(Q) = a + bQ^c$, where $Q = \sum q_i$. Making the appropriate assumptions on the parameters a, b, and c ensures that this is a downward-sloping curve, as sketched in Fig. 13.2. Hence, the more of the commodity is supplied to the market, the lower the resulting market price P will be. We assume that the production costs are such that there are negative fixed costs K, whereas the marginal costs are k. Imagine that some firms happen to have found a well where water emerges at no cost, but each bottle costs k, and each firm gets a lump-sum subsidy from the local town council if it operates a well. Given the assumptions on costs, each firm might be willing to produce any quantity at a price greater or equal to k. But it prefers to produce the output that maximizes its profits. The parameters for the underlying economic model can be found in the appendix.

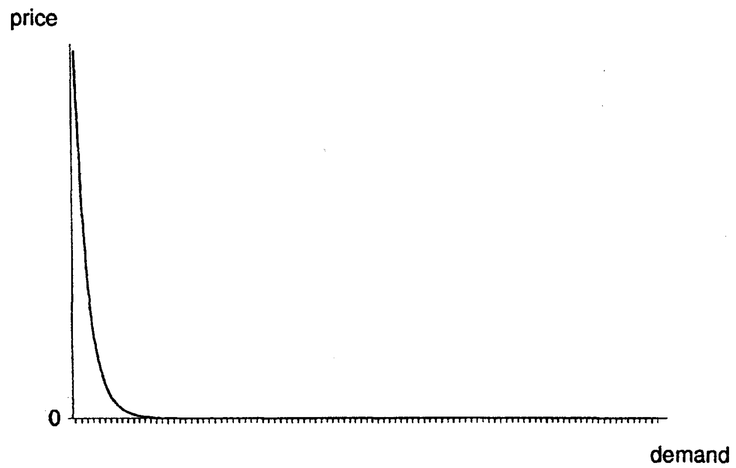

Fig. 13.2. Sketch demand function

Assume that each individual firm (there are 40 firms in our implementation) does not know what the optimal output level is, and that instead it needs to learn which output level would be good. Then, there are two basic ways to implement a GA. The first is as a model of *population learning*. Each individual firm in the population is characterized by an output rule, which is, e.g., a binary string of fixed length, specifying simply the firm's production level. In each trading day, every firm produces a quantity as determined by its output rule, the market price is determined, and the firms' profits are determined. After every 100 trading days, the population of output rules is modified by applying some reproduction, crossover, and mutation operators.[3] The underlying idea is that firms look around, and tend to imitate, and re-combine ideas of other firms that appeared to be successful. The more

[3] In each of the 100 periods between this, a firm adheres to the same output rule. This is done to match the individual learning GA (see below), and in particular its speed, as closely as possible.

successful these rules were, the more likely they are to be selected for this process of imitation and re-combination, where the measure of success is simply the profits generated by each rule. Figure 13.3 shows both the Cournot market process, and the population learning process with the GA.

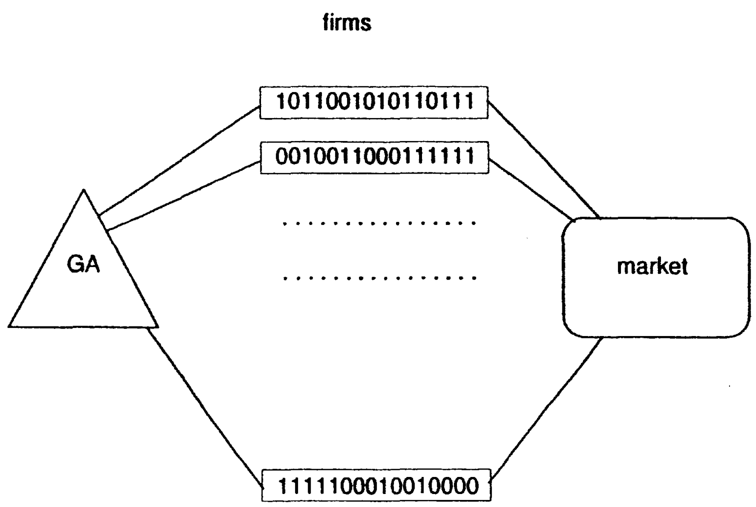

Fig. 13.3. Social learning GA

The second way to implement a GA is to use it as a model of *individual learning.* Instead of being characterized by a single output rule, each individual firm now has a set of rules in mind, where each rule is again modeled as a string, with attached to each rule a fitness measure of its strength or success, i.e., the profits generated by that rule when it was activated. Each period only one of these rules is used to determine its output level actually supplied to the market; the rules that had been more successful recently being more likely to be chosen. On top of this Classifier System, the GA, then, is used every 100 periods to modify the set of rules an individual firm has in mind in exactly the same way as it was applied to the set of rules present in the population of firms above. Hence, instead of looking how well other firms with different rules were doing, a firm now checks how well it had been doing in the past when it used these rules itself. [4] Figure 13.4 shows the underlying economic market process, and the individual learning process. [5]

Figure 13.5 presents the time series of the output levels for a representative run of each algorithm. As we see, they approach a different level. Whereas both series start around 1000, the population learning GA quickly 'converges'

[4] Hence, an alternative way to obtain the match in the speed of learning of the individual and the social learning GA would have been to endow the individual learning GA with the capability to reason about the payoff consequences for every possible output level in its set, updating all strengths every period.

[5] The parameter specification of the GA can be found in the appendix, and the pseudo-code in [12].

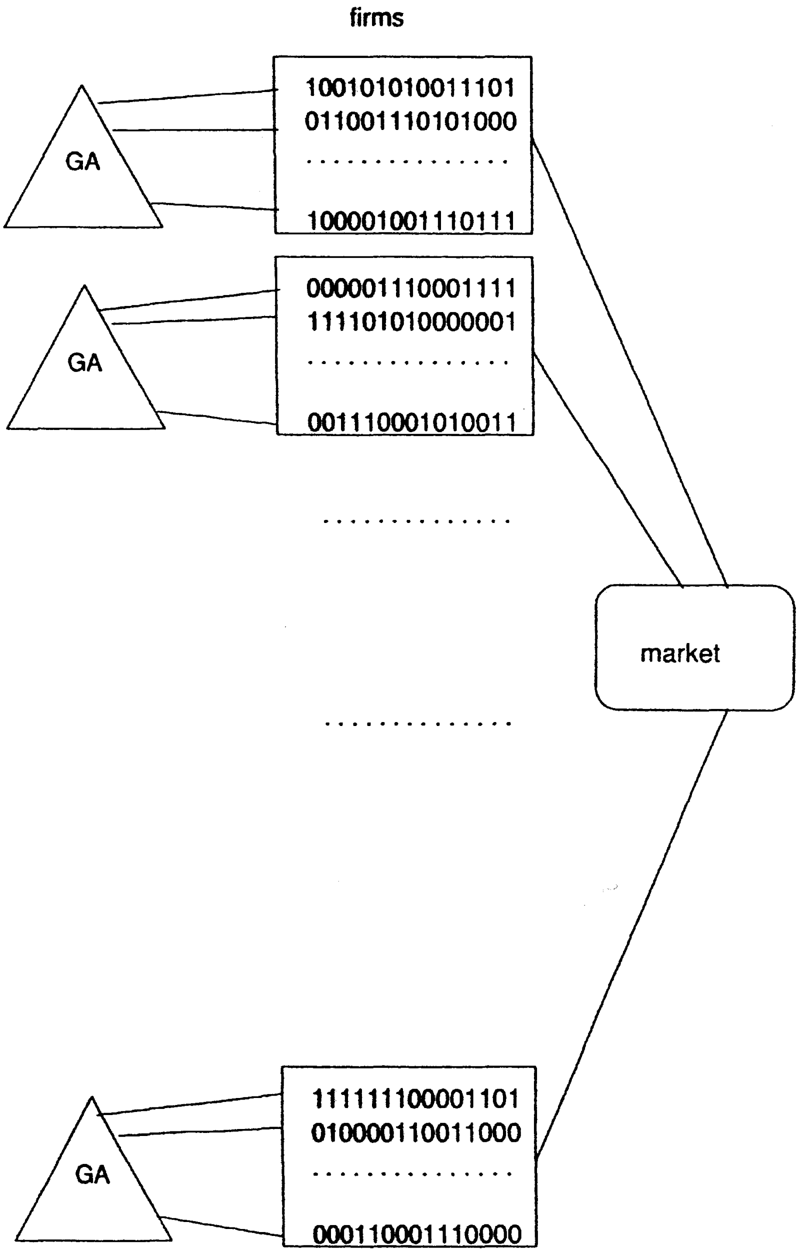

Fig. 13.4. Individual learning GA

to a level of 2000, but the individual learning GA keeps moving around a level just below 1000. [6] We want to stress that these data are generated by exactly the same identical GA for exactly the same identical underlying economic model.

[6] The 5,000 periods here presented combined with the GA rate of 100 imply that the GA has generated 50 times a new generation in each run. Each single observation in a given run is the average output level for that generation. We did all runs for at least 10,000 and some up to 250,000 periods, but this did not add new developments.

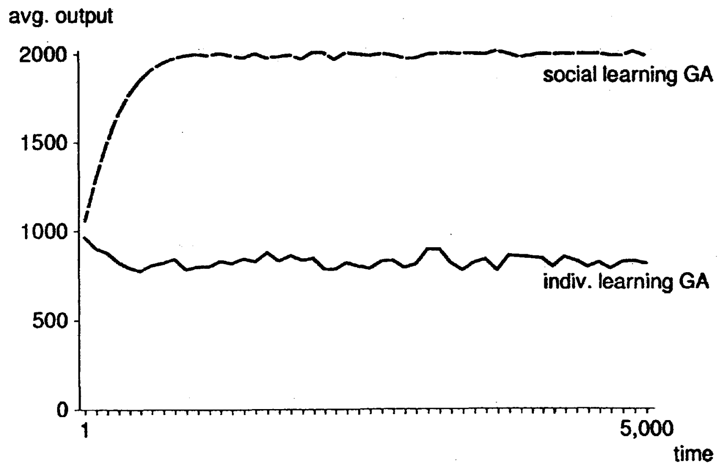

Fig. 13.5. Output levels individual learning GA and social learning GA

13.3 Analysis

We first present two equilibria of the static Cournot oligopoly game specified above for the case in which the players have complete information. The GAs do not use this information, but the equilibria serve as a theoretical benchmark that helps us understanding what is going on in the GAs. [7]

If the firms behave as price-takers in a competitive market, they simply produce up to the point where their marginal costs are equal to the market price P. Given the specification of the oligopoly model above, this implies an aggregate output level of $Q^w = 80,242.1$, and in case of symmetry, an individual Walrasian output level of $\frac{Q^w}{n} = 2,006.1$. If, instead, the firms realize that they influence the market price through their own output, still believing that their choice of q does not directly affect the output choices of the other firms, they produce up to the point where their marginal costs are equal to their marginal revenue. This leads to an aggregate Cournot-Nash equilibrium output of $Q^N = 39,928.1$, and with symmetry to an individual Cournot-Nash output of $\frac{Q^N}{n} = 998.2$.

As we see in Fig. 13.5, the GA with individual learning moves close to the Cournot-Nash output level, whereas the GA with population learning 'converges' to the competitive Walrasian output level. The explanation for this is the spite effect.

In order to give the intuition behind the spite effect in this Cournot game, let us consider a simplified version of a Cournot duopoly in which the inverse demand function is $P = a + bQ$, and in which both fixed and marginal costs

[7] The formal derivation of the two equilibria can be found in [12].

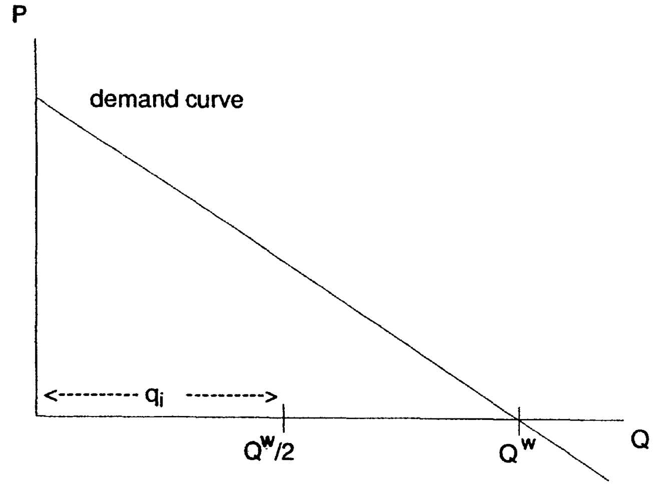

Fig. 13.6. Example simple Cournot duopoly

are zero (see [8]). The Walrasian equilibrium is then $Q^W = -a/b$, as indicated in Fig. 13.6. Suppose firm i produces its equal share of the Walrasian output: $q_i = Q^W/2$. If firm j would do the same, aggregate output is Q^W, the market price P will be zero, and both make a zero profit. What happens when firm j produces more than $Q^W/2$? The price P will become negative, and both firms will make losses. But it is firm i that makes less losses, because it has a lower output level sold at the same market price P. What happens instead if firm j produces less than $Q^W/2$? The price P will be positive, and hence this will increase firm j's profits. But again it is firm i that makes a greater profit, because it has a higher output level sold at the same market price P. In some sense, firm i is free riding on firm j's production restraint. Hence, in this Cournot duopoly the firm that produces its equal share of Q^W will always have the highest profits.

How do these payoff consequences due to the spite effect explain the difference in the results generated by the two GAs? As we saw above, the spite effect is a feature of the underlying Cournot model, and is independent from the type of learning applied. The question, then, is how this spite effect is going to influence what the firms learn. It turns out, this depends on how the firms learn.

In the population learning GA, each firm is characterized by its own production rule (see Fig. 13.3). The higher a firm's profits, the more likely is its production rule to be selected for reproduction. Due to the spite effect, whenever aggregate output is below Walras, this happen to be those firms that produce at the higher output levels. And whenever aggregate output is above Walras, the firms producing at the lowest output are most likely to be selected for reproduction. As a result, the population of firms tends to converge to the Walrasian output.[8]

[8] This is not due to the specifics of this simple example, but it is true with great generality in Cournot games (see [11], and [7]). In particular, it also holds when

In the individual learning GA, however, the production rules that compete with each other in the learning process do not interact with each other in the same Cournot market, because in any given period, an individual firm actually applies only one of its production rules (see Fig. 13.4). Hence, the spite effect, while still present in the market, does not affect the learning process, since the payoff generated by that rule is not influenced by the production rules that are used in other periods. Clearly, there is a spite effect on the payoffs realized by the other firms' production rules, but those do not compete with this individual firm's production rules in the individual learning process.

We would like to stress that it is these learning processes that is the crucial feature here, and not the objectives of the agents. Both the individual and the population learners only try to improve their own absolute payoffs. The only difference is that their learning is based on a different set of observations. The direct consequence is that in the social learning GA the spite effect drives down the performance level, while the performance of the individual learning GA improves over time. In other words, the dynamics of learning and the dynamics of the economic forces as such interact in a different way with each other in the two variants of the GA, and this explains the very different performance of the two GAs. [9]

13.4 Discussion

Before we draw some general conclusions, let us discuss some specific issues in order to put our example into a broader perspective. First, the spite effect we presented occurs in finite populations where the agents 'play the field'. The finite population size allows an individual agent to exercise some power,

all firms start at the Cournot-Nash equilibrium (as long as there is some noise in the system).

[9] There is one additional issue to be analyzed. As we saw above in Fig. 13.5, "convergence" with the individual learning GA is not as neat as with the population learning GA. Some numerical analysis shows that this is not a flaw of the individual learning GA, but related to the underlying economic model. The formal analysis of the Cournot model shows that there is a unique symmetric Cournot-Nash equilibrium. But in our numerical model we use a discrete version of the model, as only integer output levels are allowed. As a result, there turn out to be 1637 symmetric Cournot-Nash equilibria; for any average output level of the other firms from 1 to 1637, the best response for an individual firm is to choose exactly the same level. Hence, the outcomes of the individual learning GA are determined by the underlying economic forces, but convergence can take place at any of these Cournot-Nash equilibrium levels. As a result, the output levels actually observed in the individual learning GA depend in part on chance factors such as the initial output levels, the length of the bit string, and genetic drift. Notice that although there are multiple Cournot-Nash equilibria, they are all still distinct from the Walrasian equilibrium.

and influence the outcomes of the other agents. For example, in a Cournot model, when the population size n approaches infinity, the Cournot-Nash output level converges to the competitive Walrasian output. Finite populations are typically the case in GAs. To see why the 'playing the field' aspect is important, suppose there are many separate markets for different commodities, such that the actions in one market do not influence the outcomes in other markets, whereas firms can learn from the actions and outcomes in other markets. Since the spite effect does not cross market boundaries, if all firms in one market produce at the Cournot-Nash level, they will realize higher profits than the firms in another market producing at a Walrasian level. 'Playing the field' is typically the case in, e.g., economic models where the agents are firms competing in the same market. There are also some results concerning the spite effect with respect to, e.g., 2-person games in infinite populations with agents learning about the results in other games, but the occurrence of a spite effect becomes a more complicated matter (see [6]). Notice, however, that with an individual learning GA a spite effect can never occur.

Second, although, as we have seen, the spite effect may influence the outcomes of a coevolutionary process, one should not confuse the spite effect with the phenomenon of coevolution as such. In fact, as the bimatrix game in the introduction showed, the spite effect can occur in a static, one-period game, and is intrinsically unrelated to evolutionary considerations.

Third, the simple Cournot model we considered is not a typical search problem for a GA; not even if the demand and cost functions are unknown. The appeal of the Cournot model is not only that it is convenient for the presentation because it is a classic discussed in every microeconomics textbook, but the fact that we can derive formally two equilibria provided us also with two useful benchmarks for the analysis of the outcomes generated by the Genetic Algorithms. Hence, the Cournot model is just a vehicle to explain the point about the essential difference between individual and population learning GAs, and for any model, no matter how complicated, in which a spite effect occurs this essential difference will be relevant.

Fourth, one could consider more complicated strategies than the simple output decisions modeled here. For example, the Cournot game would allow for collusive behavior. However, as is well-known from the experimental oligopoly literature, dynamic strategies based on punishment and the building up of a reputation are difficult to play with more than two players. Moreover, considering more sophisticated dynamic strategies would merely obscure our point, and there exists already a large literature, for example, on GAs in Iterated Prisoners' Dilemma (see. e.g., [1], [5], or [10]).

Fifth, we are sure that the GAs we have used are too simple, and that much better variants are possible. However, bells and whistles are not essential for our point. The only essential aspect is the level at which the learning process is modeled, and the effect this has on the convergence level.

The general conclusion, then, is the following. We showed that the presence of the spite effect implies that there is an essential difference between an individual learning and a population learning GA. On the one hand, this means that when interpreting outcomes of a GA, one needs to check which variant is used, and one needs to check whether a spite effect driving the results might be present. On the other hand, it also has implications for the choice of the learning type of a GA. If a GA is used to model behavior in the social sciences, it seems ultimately an empirical question whether people tend to learn individually or socially. But if, instead, the GA is used to 'solve' some search problem, the presence of a spite effect implies that the performance of a population learning GA will be severely hindered.

A Appendix

Table 13.1. Parameters Cournot oligopoly model

inverse demand function	$P(Q)$	$a + b \cdot Q^c$
demand parameter	a	$-1 \cdot 10^{-}97$
demand parameter	b	$1.5 \cdot 10^{95}$
demand parameter	c	-39.999999997
fixed production costs	K	$-4.097 \cdot 10^{-94}$
marginal production costs	k	0
number of firms	n	40

Table 13.2. Parameters genetic algorithm

minimum individual output level	1
maximum individual output level	2048
encoding of bit string	standard binary
length of bit string	11
number rules individual GA	40
number rules population GA	40·1
GA-rate	100
number new rules	10
selection	tournament
prob. selection	fitness/$\sum$fitness
crossover	point
prob. crossover	0.95
prob. mutation	0.001

References

1. Axelrod R. (1987) The Evolution of Strategies in the Iterated Prisoner's Dilemma. In: Davis L. (Ed.), Genetic Algorithms and Simulated Annealing. London: Pitman
2. Fouraker L. E., Siegel S. (1963) Bargaining Behavior. New York: McGraw-Hill
3. Hamilton W. (1970) Selfish and Spiteful Behavior in an Evolutionary Model. Nature **228**, 1218–1225
4. Levine D. K. (1998) Modeling Altruism and Spitefulness in Experiments. Review of Economic Dynamics **1**
5. Miller J. H. (1996) The Coevolution of Automata in the Repeated Prisoner's Dilemma. Journal of Economic Behavior and Organization **29**, 87–112
6. Palomino F. (1995) On the Survival of Strictly Dominated Strategies in Large Populations. Institute for Economic Analysis WP 308.95. Universitat Autonoma de Barcelona
7. Rhode P., Stegeman M. (1995) Evolution through Imitation (with Applications to Duopoly). Working Paper E95-43. Virginia Polytechnic Institute & State University, Department of Economics
8. Schaffer M. E. (1988) Evolutionary Stable Strategies for a Finite Population and a Variable Contest Size. Journal of Theoretical Biology **132**, 469–478
9. Schaffer M. E. (1989) Are Profit Maximizers the Best Survivors? Journal of Economic Behavior and Organization **12**, 29–45
10. Stanley E. A., Ashlock D., Tesfatsion L. (1994) Iterated Prisoner's Dilemma with Choice and Refusal of Partners. In: Langton C. G. (Ed.) Artificial Life III. Redwood, CA: Addison-Wesley, 131–175
11. Vega-Redondo F. (1997) The Evolution of Walrasian Behavior. Econometrica **65**, 375–384
12. Vriend N. J. (2000) An Illustration of the Essential Difference between Individual and Social Learning, and its Consequences for Computational Analyses. Journal of Economic Dynamics and Control **24**, 1–19